3К

УДК 004.052.2

Е.Б. Козеренко, Н.В. Лунева, И.В. Галина, Ю.И. Морозова
Институт проблем информатики РАН, г. Москва, Россия
kozerenko@mail.ru, nl2@mail.ru, irn_gl@mail.ru

Лингвистические и металингвистические представления в интеллектуальных многоязычных системах

В данной работе предложен функционально-семантический подход, который обеспечивает синергетическое сочетание статистических методов и логико-лингвистических правил. В процессе грамматического разбора используются значения вероятности каждого узла разбора предложения. Взаимодействие функциональных блоков и подсистем интеллектуальной многоязычной системы между собой, а также взаимодействие ее с пользователем организуется с помощью метаданных управления и связи. Были изучены структуры когнитивного переноса в рамках поля функционального переноса первичной и вторичной предикации для русско-французской языковой пары по аналогии с русско-английской языковой парой. Материалом анализа послужили параллельные тексты, статьи из научной периодики.

1 Актуальные направления лингвистического моделирования для систем машинного перевода и обработки знаний

Современный период развития исследований и разработок в области машинного перевода и систем извлечения знаний из текстов характеризуется интенсивным процессом «гибридизации» подходов и моделей. Потребность в этом носит объективный характер. Значительные вычислительные ресурсы современных систем позволяют накапливать и использовать ранее переведенные текстовые фрагменты, обеспечивать машинный перевод, основанный на прецедентах («Example-Based Machine Translation») [1-3], эффективно поддерживать компоненту «переводческой памяти» («Translation Memory»).

Создатели систем, основанных на правилах, вводят в правила различные стохастические модели, которые позволяют отобразить динамику и разнообразие языковых форм и значений, порождаемых в процессе речевой деятельности, а сторонники статистических методов построения лингвистических моделей все чаще обращаются к подходам, основанным на лингвистических знаниях, рассматривая это как средства «интеллектуализации» систем [4-8].

Для машинного перевода наиболее сложной проблемой является реализация языковых трансформаций, которые необходимо производить при переводе с одного языка на другой. Текущий этап развития систем машинного перевода характеризуется исследованиями в области когнитивной семантики, вероятностных языковых моделей и разработкой семантико-синтаксических представлений, учитывающих многозначность и неоднозначность синтаксических структур.

При создании компьютерных моделей русской грамматики доминировали подходы, основанные на грамматиках зависимостей [9] или локально-синтаксических представлениях, при этом не разрабатывались грамматики составляющих.

Целью данной работы является разработка методов компактного представления структуры составляющих предложения (грамматика фразовых структур), с одной стороны, а с другой стороны, учитывает механизмы зависимости между узлами дерева предложения.

2 Функциональный подход к лингвистическому моделированию

Разработка понятия функции, являющегося центральным в функциональной грамматике, связана с широкой проблематикой функций языка [10-18]. Функции связаны со значениями языковых единиц, но они не тождественны им. Исследование функции некоторой языковой формы включает анализ ее значения (или ряда значений в случае многозначности).

На современном этапе лингвистических исследований и разработок необходимо синергетическое сочетание функционального и уровневого подходов. Функциональный подход интегрирует разноуровневые языковые средства (синтаксические, лексические, словообразовательные и словоизменительные) на основе их функционально-семантических характеристик.

Грамматика данного типа рассматривает в единой системе средства, относящиеся к разным языковым уровням, но объединенные на основе их семантических функций; при описании языкового материала используется подход «от семантики к ее формальному выражению» («от функций к средствам») как основной, определяющий построение грамматики, в сочетании с подходом «от формы к семантике» («от средств к функции»). Под единицами строя языка подразумеваются, прежде всего, грамматические формы слова и синтаксические конструкции, а также единицы «строевой лексики» (по Л.В. Щербе) [19]: модальные и фразовые глаголы, слова типа «вчера», «обычно», «часто», «прежде», «долго» и т.п. Ю.С. Степанов [18] вводит понятие функтора как языкового средства транспозиции одного множества языковых единиц в другое множество языковых единиц того же языка. Функция есть свойство или значение функтора. Приводится пример функции: если принять за исходное множество единиц русские глаголы типа «сообщать», «выражать», «исполнять» и т.п., а за производное множество единиц «сообщение», «выражение», «исполнение» и нечто «сообщается», «исполняется», то отношение первого множества ко второму будет функцией, а языковые формы -ение, нечто -(ает)ся будут языковыми средствами этой функции, функторами. Понятие функции является одним из центральных в коммуникативной грамматике Г.А. Золотовой. Функция – это предназначенность элемента к определенному способу существования в системе, к определенному служению этой системе [20]. Функции реализуются при взаимодействии языковых объектов и их контекстов.

Для рассмотрения семантики способов конфигурирования языковых структур мы пользуемся понятием структурного знака [21], предложенным в семиотической лингвистике С.К. Шаумяном. При этом слово также рассматривается нами не с точки зрения его лексического значения, а как функционально-категориальная единица, минимальный структурный знак. Такой подход принимается нами как определенный этап исследований структурных знаков, продиктованный необходимостью максимально полного извлечения семантической информации из возможных способов конфигурирования языковых объектов и изучения когнитивных механизмов линеаризации языковых структур.

Семиотическая лингвистика вводит понятие суперпозиции функций, полагая, что каждый языковой объект обладает исходной первичной функцией, а происходящие в действующем языке сдвиги значений – это наложение вторичной и других функций на исходную. Таким образом, использование инструмента суперпозиции [2], [22] категорий дает возможность выразить функциональные свойства языковых объектов.

3 Изофункциональные трансформации при переводе

Функциональный подход, исследующий отношения «функциональной синонимии» разнородных и разноуровневых единиц языка, чрезвычайно актуален в настоящий момент, когда проводятся эксперименты по выявлению изофункциональных и изосемичных языковых структур из параллельных текстовых корпусов. Именно этот подход позволяет найти соответствия в текстах на разных языках. В самом деле, заранее нельзя с полной достоверностью определить, каким именно образом была переведена та или иная языковая структура в текстовом корпусе. Поэтому необходимо строить и исследовать различные гипотезы при проектировании лингвистического процессора.

Отсутствие полного совпадения между английскими и русскими языковыми конструкциями в научно-технических текстах можно обнаружить при изучении сравнительной частоты употребления в них отдельных частей речи, что важно для построения систем перевода, использующих машинное обучение.

Для научного изложения в целом характерен признак номинативности, т.е. более широкое использование существительных, чем в других функциональных стилях. При этом сопоставительный анализ переводов показывает, что в русском языке эта тенденция выражена более четко, и при переводе английские глаголы нередко заменяются существительными. Проведенные нами статистические исследования параллельных текстов позволяют сделать вывод о том, что русский текст приблизительно на 35% более номинативен, чем английский. Рассмотрим следующие примеры глагольно-именных трансформаций при англо-русском переводе.

1) *The fuel system is designed to store liquid gasoline and to deliver it to the engine cylinders in the form of vapor mixed with air.*

Система питания предназначается для заправки жидким топливом и подачи его в цилиндры в виде смеси паров бензина с воздухом.

to store and to deliver → для заправки и подачи

2) *A similar approach has marked the EU's efforts to expand the current club of 15 countries to embrace former communist countries further east.*

Точно таким же подходом характеризуются усилия ЕС по расширению нынешнего клуба 15 стран дальше на восток путем присоединения к нему бывших коммунистических стран.

to embrace → по расширению

Нами были проведены исследования на материале имеющихся в нашем распоряжении параллельных переводов научных статей и отдельно взятых примеров высказываний с исследуемыми конструкциями, а также мы обращались к опросу экспертов-переводчиков. Наиболее продуктивные типы глагольно-именных трансформаций при англо-русском переводе коррелируют со следующими функциональными значениями.

– Обстоятельства цели и следствия, выраженные инфинитивом (58% в письменных переводах и 71% при опросе респондентов – профессиональных переводчиков).

3) *In order to understand the phenomenon one should consider the laws of motion.*

Для понимания этого явления надо рассмотреть законы движения.

In order to understand → Для понимания

– Составное сказуемое с инфинитивом (be + инфинитив) (51% в письменных переводах, 59% при опросе респондентов – профессиональных переводчиков).

4) *The difficulty will be to obtain the substance in question.*

Трудность будет состоять в получении рассматриваемого вещества.

to obtain → в получении

– Инфинитив после относительных местоимений which и whom с предшествующим предлогом часто переводится отглагольным существительным с предлогом «для»; в этом случае относительное местоимение с предлогом не переводится (48% в письменных переводах и 52% при опросе респондентов – профессиональных переводчиков).

5) *In vacuum molecules have large space in which to move.*

В вакууме молекулы имеют большое пространство для движения.

in which to move→ для движения

– Адъективные трансформации инфинитива (существительное + инфинитив в определительной функции) (практически 100% в обоих случаях).

6) *The amount of polonium to be obtained from a uranium mineral can be simply calculated.*

Количество полония, которое должно быть получено из урана, можно довольно просто подсчитать.

to be obtained→ которое должно быть получено

– Инфинитив в функции второго дополнения (глаголы cause, get, lead, make + инфинитив) (42% в письменных переводах и 58% при опросе респондентов – профессиональных переводчиков).

4 Вероятностная грамматика замещения деревьев

Рассмотрим, каким образом значения вероятности используются в процессе грамматического разбора. Например, вероятностная контекстно-свободная грамматика (PCFG – Probabilistic Context Free Grammar) и вероятностная грамматика подстановки деревьев (PTSG – Probabilistic Tree Substitution Grammar) присваивают вероятность (*P*) каждому дереву разбора *T* (т.е. каждому деривату) предложения *S*. Эта информация является ключевой для разрешения неоднозначности синтаксических структур. Вероятность каждого возможного дерева разбора *T* определяется как произведение вероятностей всех правил *r*, используемых для развертывания каждого узла *n* в дереве разбора:

$$P(T,S) = \prod_{n \in T} p(r(n)) . \qquad (1)$$

Вероятность однозначного предложения (т.е. предложения, где нам не надо разрешать неоднозначность) равна вероятности единственного дерева разбора для этого предложения, т.е. *P(T, S) = P(T)*. Вероятность же неоднозначного предложения равна сумме вероятностей всех возможных деревьев разбора ($\tau(S)$) данного предложения:

$$P(S) = \sum_{T \in \tau(S)} P(T,S) = \sum_{T \in \tau(S)} P(T) . \qquad (2)$$

Вероятность полного разбора предложения вычисляется с учетом категориальной информации для каждой головной вершины каждого узла. Пусть *n* – синтаксическая категория некоторого узла *n*, а *h(n)* – головная вершина узла *n*, *m(n)* – материнский узел для узла *n*, таким образом, мы будем вычислять вероятность *p(r(n)|n, h(n))*, для этого мы преобразовываем выражение (1) таким образом, что каждое правило становится обусловленным своей головной вершиной:

$$P(T,S) = \prod_{n \in T} p(r(n) \mid n, h(n)) \times p(h(n) \mid n, h(m(n))) . \qquad (3)$$

В нашей системе грамматики функциональные значения языковых структур определяются категориальными значениями головных вершин. Вероятностные характеристики вводятся в правила унификационной грамматики в виде весов, присваиваемых деревьям

разбора. Неоднозначные и многозначные синтаксические структуры учитываются в многовариантной грамматике когнитивного трансфера (переноса). Неоднозначность является коренным свойством естественного языка и вызывает основные затруднения при создании систем машинного перевода.

5 Метаданные в многоязычной лингвистической базе знаний «Интертекст»

Функционирование интеллектуальных многоязычных систем основывается на взаимодействии ряда функциональных подсистем и модулей, обеспечивающих ввод обрабатываемого текста и его разметку, распознавание функционально-семантических структур введенного текста, построение на их основе набора соответствующих структур для результирующих текстов и формирование самих текстов, а также подбор примеров из архивов ранее переведенных фрагментов текстов. Кроме того, обеспечивается управление функционированием системы и ее настройками, служебными и пользовательскими архивами, использованием словарей, эффективные управление и поддержка компоненты «переводческой памяти» [23], [24].

Взаимодействие функциональных блоков и подсистем интеллектуальной многоязычной системы между собой, а также взаимодействие ее с пользователем организуется с помощью метаданных управления и связи. Метаданные, описывающие текущее состояние программного комплекса в процессе выполнения задания (метаописание), формируются в самом начале сеанса работы пользователя на основе шаблона либо загружаются из вызванного при входе в систему файла описания предыдущего сеанса работы. Во время сеанса метаописание модифицируется соответствующими активными модулями программного комплекса интеллектуальной системы так, чтобы отображать текущее состояние системы и проекта, что позволяет при необходимости приостановить заданный процесс обработки текста и впоследствии продолжить его с того же самого места, а также изменять текущие настройки и параметры работы [25].

Метаданные обеспечивают обмен между подсистемами и блоками информацией о текущем состоянии и выполняемых действиях, требуемых для работы внутренних данных и ресурсах, а также управление специально сформированными и накопленными коллекциями данных интеллектуальной системы. Кроме того, метаданные хранят настройки программной среды на конкретную задачу и пользователя.

Наряду с использованием метаданных для организации взаимодействия функциональных блоков и подсистем между собой и управлением всей системой, метаданные применяются для описания накопленных архивов и коллекций интеллектуальной системы и их содержимого, а также их управлением.

Данные интеллектуальной многоязычной системы можно разделить на две основные категории – данные, обрабатываемые системой, и данные, используемые функциональными подсистемами для их обработки, обмена информацией между собой, настройки системы и управления ее функционированием. В общем случае обрабатываемые данные представляют собой текстовые файлы в одном из открытых форматов текстовых документов, в частном простейшем случае это может быть предложение, введенное пользователем системы в окне ввода.

Данные, используемые интеллектуальной многоязычной системой, в основном формируются в процессе работы системы и представляют собой набор разнородных массивов информации, хранящейся в различных форматах. Определенная часть данных формируется вручную на этапе разработки системы – например, иерархические системы фра-

зовых структур, описывающих язык, или коллекция параллельных текстов научных статей и патентов на русском, английском и французском языках. Часть данных, например, коллекция параллельных текстов, допускает только дополнение информации, тогда как другая часть, например, файлы конфигурации или истории работы, регулярно расширяются, обновляются и изменяются.

Интеллектуальная многоязычная система использует следующие категории данных:

– коллекции параллельных текстов научных (статьи из различных областей науки) и патентных документов на русском, английском, французском языках в различных форматах – rtf, txt, html и др.;

– коллекции фрагментов «переводческой памяти»;

– комплексы фразовых структур, образующих иерархические системы правил для каждого обрабатываемого языка, и описание языков;

– словари (одно- и двуязычные, специализированные, тематические, общие и пользовательские);

– обрабатываемые системой тексты и их истории («проекты»);

– файлы конфигурации, профили и истории пользователей системы, описание файлового архива системы и прочие служебные файлы.

Метаданные такого набора разнородной информации образуют сложный информационный комплекс, каждая часть которого описывает свой специфический класс данных системы. Иерархическая структура метаданных, образующих данный комплекс, приведена в табл. 1.

Таблица 1 – Иерархическая структура комплекса метаданных, описывающих архивы интеллектуальной многоязычной системы

<table>
<tr><td colspan="5">Атрибуты элемента данных как объекта коллекции / архива</td></tr>
<tr><td>Атрибуты «проекта»</td><td>Атрибуты научного документа / патента</td><td>Атрибуты словаря</td><td rowspan="4">Атрибуты элемента «переводческой памяти»</td><td rowspan="4">Атрибуты комплекса фразовых структур</td></tr>
<tr><td colspan="2">Атрибуты логической структуры документа / патента</td><td rowspan="2">Атрибуты описания структуры словаря</td></tr>
<tr><td colspan="2">Атрибуты текстовой фрагментации документа / патента</td></tr>
<tr><td colspan="2">Атрибуты словного состава фрагмента текста</td><td>Атрибуты описания структуры словарной статьи</td></tr>
<tr><td colspan="5">Семантико-синтаксические атрибуты слова</td></tr>
</table>

Метаданные, описывающие лингвистическую и металингвистическую информацию в системе, удобно структурировать на следующие логические уровни – описания логической структуры документа, описания фрагментации текста документа на удобные для дальнейшей обработки блоки (например, фразы), описания словного состава фрагмента и описания семантико-синтаксической характеристики каждого слова.

Метаданные описания логической структуры документа должны полностью описывать его структуру: название, списки авторов, заголовки, разделы, абзацы, а также списки служебных полей (таких как коды, адреса, даты, место работы, тематические области и других). Метаданные описания фрагментации текста документа характеризуют разбиение самого текста на фразы или фрагменты фраз.

Метаданные описания семантико-синтаксических характеристик слова должны полностью охватывать его лексические и семантико-синтаксические свойства: часть речи, падеж, число, род, вид глагола, лицо, время, наклонение, а также признаки одушевленности, «собственности», возвратности, переходности, залога, степени сравнения, уникальности форм единственного и множественного числа и другие. Также метаданные описания семантико-синтаксических характеристик слова включают в себя атрибуты, построенные на основе разработки системы описания полей функционального переноса языка, с учетом как отношений зависимости, так и структуры составляющих, что имеет определенное сходство с проектными решениями грамматики HPSG [26], [27].

Метаданные описания словного состава фрагмента текста охватывают каждый элемент фрагмента текста с точки зрения его положения во фрагменте, длины, алфавитно-цифрового состава и возможных ролей в тексте. Данные характеристики словного элемента определяются до выявления его семантико-синтаксических характеристик. Кроме того, в качестве словного элемента фрагмента текста рассматриваются не только сами слова, но и знаки пунктуации, и любые встреченные внутри фрагмента цифровые и алфавитно-цифровые объекты, формулы, возможные сокращения и аббревиатуры. Структура записи, характеризующей словный элемент, и ее атрибуты приведены в табл. 2.

Таблица 2 – Атрибуты описания словного элемента фрагмента текста и их значения

атрибут	Описание
позиция	Порядковая позиция словного элемента внутри фрагмента текста.
длина	Указывает длину слова в знаках.
тип	Указывает тип слова – алфавитное, с дефисом в составе, цифровое, знак препинания, кавычки, скобки, цифро-буквенное или буквенно-цифровое, сокращение, начинающееся с заглавной буквы, состоящее из заглавных букв, иноязычное и т.п.
роль	Указывает предполагаемую роль данного словного элемента – сокращение, аббревиатура, название, имя собственное, формула и т.д.

Метаданные интеллектуальной многоязычной системы независимо от своего функционального значения и использования организованы в единую базу метаданных, включенную в ядро системы. В базе метаданных хранятся вместе с атрибутами метаданных их значения и описания, а также описание группировки метаданных. Функциональные подсистемы и модули в процессе обработки текста извлекают из базы метаданных атрибуты и значения только используемых в данный момент функциональных групп метаданных.

Ведение базы метаданных осуществляется с помощью специальной подсистемы ядра интеллектуальной многоязычной системы, доступной только специально уполномоченному пользователю. Таким образом, при необходимости обновления или расширения пространства метаданных, например добавления в систему нового языка или объекта, достаточно просто добавить в базу метаданных массив необходимых атрибутов, всесторонне описывающий вводимый объект.

6 Трансфер пропозиционального ядра в русско-французской языковой паре

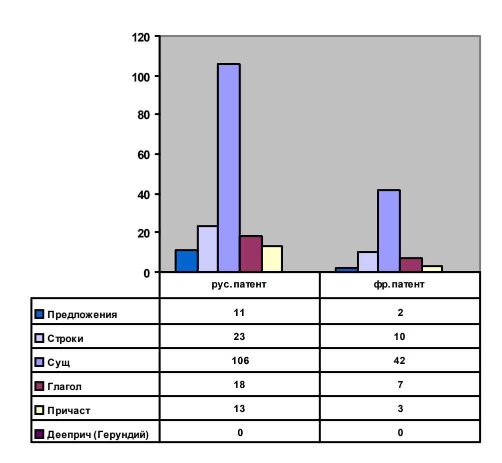
Основу семантико-синтаксической структуры предложения составляет пропозициональное ядро, прежде всего, языковые средства предикации. Были изучены структуры когнитивного переноса в рамках поля функционального переноса (ПФП) первичной и

вторичной предикации для русско-французской языковой пары по аналогии с русско-английской языковой парой. Были выделены базовые правила когнитивного переноса для различных функциональных значений (частотные характеристики были выделены на основании анализа патентных текстов). Материалом анализа послужили параллельные тексты патентов и/или рефератов патентов на русском и французском языках, взятые из базы данных Роспатента.

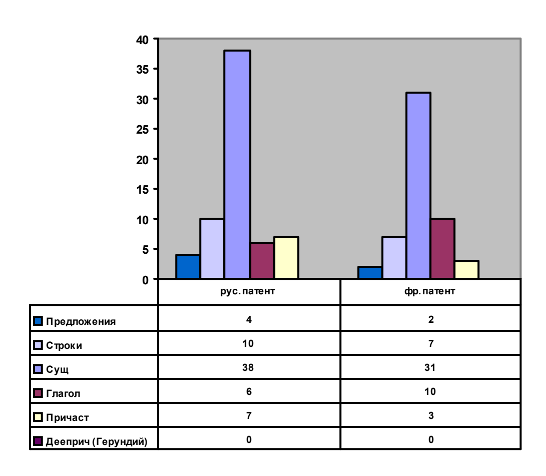
Сравнение русских и французских текстов рефератов научных патентов показало, что доля действительно параллельных текстов в них составляет примерно 30%. Остальные тексты можно назвать когнитивно-сопоставимыми, причем объем русского текста может превышать объем французского на две трети. Однако распределение по частям речи в русских и французских научных текстах патентных рефератов (и самих патентов) очень близко по составу и объёму, что отражено на рис. 1. Русский текст в целом на 30 – 35% более номинативен, чем французский, в котором в поле вторичной предикации предпочтение отдаётся инфинитиву (в русском – отглагольным существительным).

Сравнение русского и французского рефератов по объему и составу частей речи

Реферат патента WO2004009333 *Реферат патента WO2004017987*



	рус.патент	фр.патент
Предложения	11	2
Строки	23	10
Сущ	106	42
Глагол	18	7
Причаст	13	3
Дееприч (Герундий)	0	0



	рус.патент	фр.патент
Предложения	4	2
Строки	10	7
Сущ	38	31
Глагол	6	10
Причаст	7	3
Дееприч (Герундий)	0	0

Рисунок 1 – Распределение по частям речи в русских и французских научных текстах патентных рефератов

В первом примере тексты рефератов не параллельные, а когнитивно-сопоставимые, во втором тексты русского и французского реферата параллельны: перевод выполнен точно, почти дословно. В любом случае, как видно из примеров, и в русских, и во французских патентных текстах существенно высока доля именных групп, что вообще всегда характерно для научно-технических текстов.

6.1 Сравнение морфологических способов выражения цели действия в русском и французском языках

Правила когнитивного переноса для функциональных значений цели и назначения представлены на следующей схеме (рис. 2).

Цель, назначение

Русский язык		*Французский язык*
Существительное (98)		**Инфинитив (72)**
Инфинитив (2)	←→	Существительное (26)
Придаточное предложение (0)		Придаточное предложение (2)

Рисунок 2 – Правила когнитивного переноса для функциональных значений цели и назначения

Таким образом, набор структур, используемых для выражения цели действия, одинаков для русского и французского языков, однако французский тяготеет к инфинитивной структуре, а русский – к именной *(Для увеличения способности сети к обобщению... / Afin d'augmenter la capacité du réseau de généraliser...).*

Примеры:.

1. [Cat : VerbNoun] *{для распознавания} {pour la reconnaissance}* – предложная группа: предлог + существительное.

2. [Cat : VerbInf] *{чтобы распознать} {afin de reconnaître}* – союз + инфинитив.

3. [Cat : Sentence] *{чтобы распознавание было эффективным} {pour que la reconnaissance soit efficace}* – придаточное предложение, присоединяемое подчинительной связью (союзом цели). При трансформации русского отглагольного существительного во французский инфинитив необходимо сделать выбор между его активной и пассивной формой. Видимо, в рамках системы автоматического перевода данный выбор лучше всего осуществляется с применением статистических данных (активный инфинитив встречается в текстах намного чаще пассивного; в анализируемых текстах французский пассивный инфинитив в качестве перевода русского отглагольного существительного встретился в 13% случаев).

7 Сравнение морфологических способов выражения предшествования / следования в русском и французском языках

Данные по количественному распределению случаев использования различных морфологических способов выражения предшествования / следования для русского и французского языков, полученные при анализе патентных текстов приведены на рис. 3.

Предшествование / следование

Русский язык		*Французский язык*
Существительное (89)		**Существительное (90)**
Придаточное предложение		Инфинитив (10)
Инфинитив (2)	←→	Придаточное предложение (0)
Деепричастие (0)		Причастие (0)

Рисунок 3 – Количественное распределение случаев использования различных морфологических способов выражения предшествования / следования

Примеры:

1) *перед\после* + NP (существительное) – предложная группа (*перед* + существительное в творительном падеже): *…мешки, <u>перед использованием</u> порошка, например, для изготовления бетона.*

2) *avant\après* + NP (существительное): *<u>Avant utilisation</u> dans la réaction, le métal contenu dans le catalyseur doit être réduit.*

В патентных текстах частотность конструкций вида *предлог [перед (avant)\ после (après)] + существительное* в обоих языках практически одинакова; существенно реже придаточные предложения (в русском); французский язык придаточному предложению предпочитает инфинитивный оборот.

7.1 Таксисные значения, реализуемые в полипредикативных конструкциях

Самые частотные правила когнитивного переноса в анализируемых текстах следующие: французское придаточное относительное предложение вида (S{S1{VP+NP}+ +S2{Pron_relat{qui}+NP+VP}}) русское причастие. Наиболее продуктивные типы конструктивных трансформаций при французско-русском переводе коррелируют со следующими аспектуально-таксисными значениями [28]:

– *таксис одновременности* с основным глаголом на линейной цепочке сложного предложения, функционально соответствует participe présent: *Une filière regroupe toutes les activités productrices* <u>*qui concourent*</u> *à la fabrication d'un bien... / Производственно-технологическая цепочка объединяет все этапы производственной деятельности,* <u>*связанные с*</u> *изготовлением какого-либо продукта...* Конечно, в этой ситуации достаточно часты и совпадения конструкций (около 35%) – *фр.*: Cnj(*qui*) + V(fin) – *рус.*: Cnj(*который*) + V(fin) : *Dans d'autres gels constitués d'un mélange d'hydroxypropylcellulose et de silicate d'aluminium et de magnésium, on observe également une chute de viscosité au cours du temps* <u>*qui entraîne une sédimentation des actifs en suspension*</u> *et une hétérogénéité de la- dispersion dans le produit fini. / В других гелях, содержащих смесь гидроксипропилцеллюлозы и силиката алюминия и магния, также со временем отмечают понижение вязкости,* <u>*которое вызывает осаждение активных веществ, находящихся в суспензии*</u>*, и неоднородность дисперсии в конечном продукте*;

– *таксис предшествования и завершённости* (исчерпанности) внутреннего времени действия – французские полипредикативные конструкции с инфинитивом прошедшего времени (infinitif passé) вида (S{S1{Inf_pass{Conj{après, dès que, aussitôt que}+Inf_pass}}+ +S2{NP+VP}}); выражают предшествование на линейной цепочке, а в независимых предложениях и на оси времени). На русский язык может переводиться деепричастием (чаще всего) либо причастием или придаточным относительным предложением. Тут реализуется ограничительная функция временных союзов, задающих ограничение действия определённым пределом, и накладывается семантика этих союзов, «подкрепляя» грамматически присущее инфинитиву прошедшего времени значение предшествующего действия: <u>*Après avoir reçu de*</u> *sa banque un avis d'impayé... /* <u>*Получив от*</u> *своего банка извещение об отказе в оплате...*

7.2 Особенности перевода французского абсолютного причастного оборота

Французский абсолютный причастный оборот – Proposition participe absolue (может быть выражен формами participe présent – непредшествование, обычно одновременность; participe passé, participe passé composé – предшествование) – выступает как придаточное предложение, у которого имеется собственное подлежащее, отличное от подлежащего главного предложения: (S{S1{Proposition_part_abs{NP+Part_pr OR NP+Part_pass OR NP + +Part_pass_comp}}, S2{NP+VP}}). Данный оборот не имеет аналогов в русском языке. Чаще всего выражает отношения причины или временные отношения. Реализует следующие значения.

– *Причина и одновременность.* Обороты со значением причины обычно переводятся на русский придаточными с союзами так как, поскольку, потому что; когда, после того, как или, реже, обстоятельствами или деепричастиями: *Le travail au demeurant était d'une extrême simplicité, la machine étant presque entièrement automatique. / Поначалу работа была чрезвычайно проста, так как (потому что) станок действовал почти автоматически.* Отметим, что это предложение можно перевести на русский язык иначе: *Когда станок действовал почти автоматически, работа была чрезвычайно проста*. Но в таком переводе семантика причинной обусловленности скрыта, хотя таксис (одновременность) не нарушен.

– *Причина и предшествование.* Временные отношения в русском языке выражаются с помощью союзов когда; как только; тотчас, после того как; едва только; сразу после того как… Во французских предложениях абсолютный причастный оборот может не иметь лексически выраженных временных показателей; тем не менее, часто добавляются наречия со значением немедленного предшествования: *aussitôt, sitôt, à peine, une fois. Например: Une*

ardeur de travail l'ayant soudain ressaisi, il se rassit devant sa table. / [Поскольку] его внезапно охватила жажда работы, он снова уселся за стол. Тут французский оборот выступает с основным для него аспектуально-таксисным значением предшествования, усиленным семантикой наречия *soudain* «внезапно».

– *Комитативное значение (дополнительного действия),* при переводе используются союзы причём, также, таким образом, и: *Dans la pratique cela devrait se traduire par 250 représentation par an, le public ayant le choix... / На практике это должно будет означать 250 представлений в год, причём зрители смогут выбирать...*

– *Временные отношения: a) предшествования: Le travail ayant terminé, il se retourna... / После того, как работа была завершена, он вернулся... b) одновременности: Le protocol se lisant, le système travaille... / Пока считывается протокол, система работает...*

Абсолютный причастный оборот представляет большие трудности для машинного перевода как по причине отсутствия подобной синтаксической конструкции в русском языке, что вынуждает прибегать к трансформациям, так и по причине «диффузности» значения самой французской конструкции, особенно, если отсутствуют эксплицитно выраженные маркеры временных отношений.

Заключение

Современный этап исследований в области проектирования систем обработки естественного языка ставит задачу оптимального сочетания сильных сторон двух исследовательских парадигм: логико-лингвистического моделирования, использующего правила, и стохастического подхода.

В данной статье рассмотрены вопросы создания нового системного подхода к извлечению лингвистических и предметных знаний из параллельных текстовых корпусов на нескольких европейских языках (русском, английском, французском), позволяющего осуществить семантическое выравнивание значимых текстовых сегментов, для решения задач машинного перевода и обработки текстовых знаний. Предлагаемый подход основан на функционально-семантических представлениях языковых структур и статистических методах.

В рамках научно-исследовательской работы было проведено семантическое выравнивание фокусной группы параллельных текстов, сформированы фокусные обучающие выборки и проведен сбор статистических данных о типах изосемичных языковых структур в русском, английском и французском, немецком и итальянском дискурсе.

Рассмотрены статистические методы анализа языковых объектов, сформирован аппарат вероятностных расширений логико-лингвистических правил и выработаны методы создания алгоритмов разбора и трансфера языковых структур, сформированы статистические портреты семантических групп языковых явлений (полей функционального переноса), выявлены наиболее частотные соответствия в пределах семантических групп для каждого из рассматриваемых языков.

Литература

1. Brown R.D. Example-based machine translation in the Pangloss system / R.D. Brown // In COLING-96. – Copenhagen, 1996. – P. 169-174.
2. Козеренко Е.Б. Логико-статистические методы представления языковых структур в машинном переводе / Е.Б. Козеренко // Труды Международной конференции Диалог'2005 «Компьютерная лингвистика и интеллектуальные технологии». – М. : Наука, 2005.
3. Knight K. Automating knowledge acquisition for machine translation / K. Knight // AI Magazine. – 1997. – № 18. – P. 81-96.
4. Compendium of Translation Software / John Hutchins, editor. – [Seventh edition]. – August, 2003.

5. Lagoudaki E. Translation Memories Survey 2006: User's Perceptions Around TM Usage / E. Lagoudaki // In Proceedings of the Translating and the Computer 28 Conference, (16 – 17 November 2006, London). – Aslib/IMI, London, 2006. – P. 1-29.
6. Wang Ye-Yi. Modelling with structures in statistical machine translation / Wang Ye-Yi and Alex Waibel // In ACL 36/COLING 17. – 1998. – P. 1357-1363.
7. Alshawi H. A comparison of head transducers and transfer for a limited domain translation application / Alshawi H., Adam L. Buchsbaum and Fei Xia // In ACL 35/EACL 8. – 1997. – P. 360-365.
8. Plana E. SIMILIS Second-Generation Translation Memory Software / E. Plana // In Proceedings of the Translating and the Computer 28 Conference, (25 – 24 November 2005, London). – Aslib/IMI, London, 2005.
9. Mel'cuk Igor Aleksandrovich. Dependency Syntax: theory and practice / Mel'cuk Igor Aleksandrovich. – Albany : State University of New York, 1988.
10. Якобсон Р.О. Разработка целевой модели языка в европейской лингвистике в период между двумя войнами / Р.О. Якобсон // Новое в лингвистике. – М., 1965. – Вып. 4. – С. 372-377.
11. Якобсон Р.О. Шифтеры, глагольные категории и русский глагол / Р.О. Якобсон // Принципы типологического анализа языков различного строя. – М., 1972.
12. Степанов Ю.С. Имена, предикаты, предложения (семиологическая грамматика) / Степанов Ю.С. – М. : Едиториал УРСС, 2004.
13. Halliday M.A.K. An Introduction to Functional Grammar / Halliday M.A.K. ; ed. by Edward Arnold. – London, 1985.
14. Halliday M.A.K. System and Function in Language / Halliday M.A.K. – London : Selected Papers, 1976. – 250 p.
15. Слюсарева Н.А. Проблемы функционального синтаксиса современного английского языка / Слюсарева Н.А. – М., 1981. – 206 с.
16. Шведова Н.Ю. Один из возможных путей построения функциональной грамматики русского языка // Проблемы функциональной грамматики / Н.Ю. Шведова. – М., 1985. – С. 30-37.
17. Звегинцев В.А. Функция и цель в лингвистической теории // Проблемы теоретической и экспериментальной лингвистики / В.А. Звегинцев. – М., 1977. – С. 120-146.
18. Гак В.Г. К типологии функциональных подходов к изучению языка // Проблемы функциональной грамматики / В.Г. Гак. – М., 1985. – С. 5-15.
19. Щерба Л.В. Языковая система и речевая деятельность / Щерба Л.В. – Л., 1974. – 428 с.
20. Золотова Г.А. Коммуникативная грамматика русского языка / Золотова Г.А., Онипенко Н.К., Сидорова М.Ю. – М., 2004.
21. Шаумян С.К. Семиотическая Лингвистика как Объяснительная Наука / С.К. Шаумян // Труды Международной конференции Диалог'2005 «Компьютерная лингвистика и интеллектуальные технологии». – М. : Наука, 2005. – С. 507-513.
22. Shaumyan S. Signs, Mind, and Reality / Shaumyan S. – John Benjamins Publishing Company, USA, 2006.
23. Лунева Н.В. Архитектура и метаданные многоязычной лингвистической базы знаний / Н.В. Лунева // Системы и средства информатики. – Москва : ИПРАН, 2007. – Вып. 17. – С. 317-336.
24. Козеренко Е.Б. Проектирование многоязычного лингвистического ресурса для систем машинного перевода и обработки знаний / Е.Б. Козеренко, Н.В. Лунева, Ю.И. Морозова, П.В. Ермаков // Системы и средства информатики. – Москва : ИПРАН, 2009. – Вып. 19. – С. 119-141.
25. Лунева Н.В. Метаданные в системе управления многоязычной лингвистической базой знаний / Н.В. Лунева // Труды 10-й Всероссийской научной конференции RCDL-2008 «Электронные библиотеки: перспективные методы и технологии, электронные коллекции», (Дубна, 2008). – С. 70-75.
26. Head-Driven Phrase Structure Grammar server. – Режим доступа : http://www.ling.ohio-state.edu/research/hpsg/
27. Head-Driven Phrase Structure Grammar. – Режим доступа : http://hpsg.stanford.edu/ideas.html
28. Галина И.В. Функционально-синонимичные способы выражения аспектуально-таксисных значений во французском и русском языках (для многоязычного лингвистического процессора) / И.В. Галина // Системы и средства информатики. – М. : Наука, 2007. – Вып. 17. – С. 293-316.

Literatura

1. Brown, R.D. Example-based machine translation in the Pangloss system. In COLING-96. Copenhagen. 1996. P. 169-174.
2. Kozerenko E.B. Trudy Mezhdunarodnoj konferencii Dialog'2005 "Komp'juternaja lingvistika i intellektual'nye tehnologii". M.: Nauka. 2005.
3. Knight K. AI Magazine 18. 1997. P. 81-96.

4. Compendium of Translation Software. John Hutchins, editor, Seventh edition. August, 2003.
5. Lagoudaki, E. Translating and the Computer 28 Conference. 16-17 November 2006. Aslib/IMI, London. 2006. P. 1-29.
6. Wang Ye-Yi. ACL 36/COLING 17. 1998. P. 1357-1363.
7. Alshawi H. ACL 35/EACL 8. 1997. P. 360-365.
8. Plana E. Translating and the Computer 28 Conference. 25-24 November 2005. Aslib/IMI, London. 2005
9. Mel'cuk I. A. Dependency Syntax: theory and practice. Albany: State University of New York. 1988
10. Jakobson R.O. Novoe v lingvistike. M. Vyp 4. 1965 S. 372-377.
11. Jakobson R.O. Principy tipologicheskogo analiza jazykov razlichnogo stroja. M. 1972.
12. Stepanov Ju.S. Imena, predikaty, predlozhenija (semiologicheskaja grammatika). M.: Editorial URSS. 2004.
13. Halliday M.A.K. An Introduction to Functional Grammar. London. Ed. by Edward Arnold. 1985.
14. Halliday M.A.K. Selected Papers, London, 1976. 250 p.
15. Sljusareva N.A. Problemy funkcional'nogo sintaksisa sovremennogo anglijskogo jazyka. M. 1981. 206 s.
16. Shvedova N.Ju. Problemy funkcional'noj grammatiki. M.1985. S. 30-37.
17. Zvegincev V.A. Problemy teoreticheskoj i jeksperimental'noj lingvistiki. M. 1977. S. 120-146.
18. Gak V.G. Problemy funkcional'noj grammatiki. M. 1985. S. 5-15.
19. Shherba L.V. Jazykovaja sistema i rechevaja dejatel'nost'. L. 1974. 428 s.
20. Zolotova G.A Kommunikativnaja grammatika russkogo jazyka. M. 2004.
21. Shaumjan S.K. Trudy Mezhdunarodnoj konferencii Dialog'2005 "Komp'juternaja lingvistika i intellektual'nye tehnologii". M.: Nauka. 2005. S. 507-513.
22. Shaumyan S. Signs, Mind, and Reality. USA: John Benjamins Publishing Company. 2006.
23. Luneva N.V. Sistemy i sredstva informatiki. Vyp 17. Moskva: IPRAN. 2007. S. 317-336.
24. Kozerenko E.B. Sistemy i sredstva informatiki. Vyp 19. Moskva: IPRAN. 2009. S. 119-141.
25. Luneva N.V. Trudy 10-j Vserossijskoj nauchnoj konferencii RCDL-2008 "Jelektronnye biblioteki: perspektivnye metody i tehnologii, jelektronnye kollekcii". Dubna. 2008. S. 70-75.
26. Head-Driven Phrase Structure Grammar server: http://www.ling.ohio-state.edu/research/hpsg/
27. Head-Driven Phrase Structure Grammar: http://hpsg.stanford.edu/ideas.html
28. Galina I.V. Sistemy i sredstva informatiki. Vyp 17. M.: Nauka. 2007. S. 293-316.

Є.Б. Козеренко, Н.В. Лунєва, І.В. Галіна, Ю.І. Морозова

Лінгвістичні та металінгвістичні представлення в інтелектуальних багатомовних системах

У даній роботі запропоновано функціонально-семантичний підхід, який забезпечує синергетичне сполучення статистичних методів і логіко-лінгвістичних правил. У процесі граматичного розбору використовуються значення ймовірності кожного вузла розбору речення. Взаємодія функціональних блоків і підсистем інтелектуальної багатомовної системи між собою, а також взаємодія її з користувачем організується за допомогою метаданих керування і зв'язку. Були вивчені структури когнітивного переносу у межах поля функціонального переносу первинної і вторинної предикації для російсько-французької мовної пари за аналогією з російсько-англійською мовною парою. Матеріалом аналізу послужили паралельні тексти, статті з наукової періодики.

E.B. Kozerenko, N.V. Luneva, I.V.Galina, J.I. Morozova

Linguistic and Metalinguistic Representations in Intelligent Multilingual Systems

The approach, which integrates linguistic means of different language levels (syntactic, lexical, derivational and inflectional) on the basis of their functional semantic characteristics and ensures the synergetic combination of statistical methods and logical linguistic rules, is proposed in this work. In the process of grammatical analysis the values of probability are used. Interaction of functional blocks and subsystems of the intellectual multilingual system between themselves, and also its interaction with the user is organized with the aid of metadata of control and connection. The authors studied the structures of cognitive transfer within the framework of the functional transfer field of primary and secondary predication for the Russian-French language pair by analogy with the Russian-English language pair. The material for analysis was taken from the parallel texts of scientific periodicals.

Статья поступила в редакцию 23.06.2011.